DOCUMENT RESUME

ED 435 671 TM 030 300

AUTHOR Thompson, Bruce; Cook, Colleen
TITLE Reliability and Validity of SERVQUAL Scores Used To Evaluate Perceptions of Library Service Quality.
PUB DATE 2000-01-00
NOTE 37p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Dallas, TX, January 27-29, 2000).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Evaluation Methods; Higher Education; *Library Services; Measurement Techniques; *Reliability; Scores; Surveys; *Users (Information); *Validity
IDENTIFIERS *SERVQUAL

ABSTRACT

Research libraries are increasingly supplementing collection counts with perceptions of service quality as indices of status and productivity. The present study was undertaken to explore the reliability and validity of scores from the SERVQUAL measurement protocol (A. Parasuraman and others, 1991), which has previously been used in this type of application in libraries. The study involved collection of perceptions from 697 participants representing 4 different user groups and 3 different biannual surveys. Scores were highly reliable, but the five SERVQUAL dimensions suggested by SERVQUAL scoring keys were not recovered. Furthermore, different dimensions were recovered under three different frames of reference. (Contains 9 tables and 28 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 435 671

colandér.wp1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Bruce Thompson
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Reliability and Validity of SERVQUAL Scores Used to Evaluate Perceptions of Library Service Quality

Bruce Thompson
Texas A&M University
and
Baylor College of Medicine

Colleen Cook
Texas A&M University

TM030300

Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas, January 27-29, 2000.

Bruce Thompson is Professor of Educational Psychology and Distinguished Research Scholar, Texas A&M University, and Adjunct Professor of Community Medicine, Baylor College of Medicine; he may be contacted via Internet URL: "http://acs.tamu.edu/~bbt6147/". Colleen Cook is Professor of Library Science and Associate Dean of the Sterling C. Evans Library at Texas A&M University; she may be contacted via e-mail at address: "ccook@lib-gw.tamu.edu".

ABSTRACT

Research libraries are increasingly supplementing collection counts with perceptions of service quality as indices of status and productivity. The present study was undertaken to explore the reliability and validity of scores from the SERVQUAL measurement protocol, which has previously been used in some such applications in libraries. The study involved collection of perceptions from 697 participants representing four different user groups and three different bi-annual surveys. Scores were highly reliable, but the five SERVQUAL dimensions suggested by SERVQUAL scoring keys were not recovered. Furthermore, different dimensions were recovered under three different frames of reference.

This is an era of accountability for research libraries housed on university campuses confronting funding cutbacks and increased competition to recruit and retain tuition-paying students. In the context of our times, "every unit... is valued in proportion to its contribution to the quality success of the campus" (Nitecki, 1996b, p. 181). Traditionally, the evaluation criteria of the Association for Research Libraries (ARL) emphasized objective descriptions of collection sizes and their special features.

But more recently there has been "increasing pressure on libraries to assess the degree to which their services demonstrate criteria of 'quality.' ...The emphasis on these measures and services provided to library clientele requires librarians... not to equate 'quality' merely with collection size" (Hernon & McClure, 1990, p. xv). As Nitecki (1996b) noted, "A measure of library quality based solely on collections has become obsolete" (p. 181).

The basis for the movement beyond sole reliance on collection counts is clear. As Nitecki (1997) recently observed, "Flying across the Atlantic, are you more likely to judge the quality of the airline you use by the number of planes it operates or by the reliability of its schedules of departures and arrivals and the attention its staff gives you?" (p. 181).

Unfortunately relatively few measures have been developed that can be used to evaluate client perceptions of library service quality (Stein, 1997). As Franklin and Nitecki (1999) noted in a recent ARL white paper, "Several individual libraries have conducted independent measures of user satisfaction and characteristics of library use, but there are no systematic

reporting mechanisms for the results among research libraries" (p. 3).

Several researchers have turned to the marketing literature for a measurement protocol that can be used for this purpose. The SERVQUAL protocol, which includes 22 items ostensibly measuring perceptions of *tangibles, reliability, responsiveness, assurance* and *empathy*, has been fairly widely used for this purpose (Parasuraman, Berry & Zeithaml, 1991; Parasuraman, Zeithaml & Berry, 1985, 1994). Within this model, "only customers judge quality; all other judgments are essentially irrelevant" (Zeithaml, Parasuraman, Berry, 1990, p. 16).

The SERVQUAL scale has been described and investigated in over 100 articles and 20 doctoral dissertations (Nitecki, 1996b, p. 183). At least in the view of Andaleeb and Simmonds (1998), "Although this vein of research has been pursued with some enthusiasm, empirical support for the suggested framework and the desirability of the measurement instrument has not been very encouraging" (p. 157). Babakus and Boller (1992) present some of these criticisms. But other reports have been more favorable (cf. Nitecki, 1996a).

Nature of Reliability and Validity

It is vitally important that researchers who are investigating the psychometric properties of scores from tools measuring perceptions of library service quality understand the nature of psychometric characteristics. As Thompson (1994) observed,

> One unfortunate common feature of contemporary scholarly language is the usage of the statement,

> "the test is reliable" or "the test is valid." Such language is both incorrect and deleterious in its effects on scholarly inquiry, particularly given the pernicious consequences that unconscious paradigmatic beliefs can exact. ...Pernicious, unconscious, incorrect assumptions that tests themselves are reliable [or valid] can lead to insufficient attention to the impacts of measurement integrity on the integrity of substantive research conclusions. (p. 839-840)

For example, as Rowley (1976) argued regarding reliability, "It needs to be established that an instrument itself is neither reliable nor unreliable.... A single instrument can produce *scores* which are reliable, and other *scores* which are unreliable" (p. 53, emphasis added). Similarly, Crocker and Algina (1986, p. 144, emphasis added) argued that, "...A test is not 'reliable' or 'unreliable.' Rather, reliability is a property of the *scores* on a test for a *particular* group of examinees."

In another widely respected measurement text, Gronlund and Linn (1990, emphasis in original) noted,

> Reliability refers to the *results* obtained with an evaluation instrument and not to the instrument itself.... Thus, it is more appropriate to speak of the reliability of the "test scores" or of the "measurement" than of the "test" or the "instrument." (p. 78)

All this means that the survey respondents "themselves impact the

reliability of scores, and thus it becomes an oxymoron to speak of 'the reliability of the test' without considering to whom the test was administered, or other facets of the measurement protocol" (Thompson, 1994, p. 839). Indeed, the recognition of these realities has led to the development of the "reliability generalization" method proposed by Vacha-Haase (1998) to characterize (a) typical score reliability, (b) the variability of score reliability, and (c) the measurement features that explain or predict variation in score reliability across test administrations.

Thus, a measure such as SERVQUAL may work in industrial settings, but not libraries. Or the measure may yield useful scores on some campuses, but not on others. Or scores from one user group (e.g., faculty, graduate students) may be useful, while scores from another user group (e.g., undergraduate students) may not be.

Purpose of the Study

The present study was undertaken to address two research questions. First, how reliable are the various SERVQUAL scores across different times of measurement (1995, 1997, and 1999) and across different respondent user groups (i.e., faculty, staff, and undergraduate and graduate students)? Second, does factor analysis of SERVQUAL responses yield the structure suggested by the measure's scoring keys (i.e., factors of *tangibles, reliability, responsiveness, assurance* and *empathy*), and thus corroborate score validity?

Methods

Participants

The participants in the study were 697 faculty, staff, and

undergraduate and graduate students who completed a SERVQUAL evaluation of the main research library at a large southwestern university in 1995 (n_{95} = 179), 1997 (n_{97} = 287), and 1999 (n_{99} = 231). The participants were selected by randomly sampling from various campus databases. Table 1 provides a breakdown of the sample across both time and the user groups.

INSERT TABLE 1 ABOUT HERE.

Instrumentation

The 697 participants rated service quality of the library using the 22 SERVQUAL items. The set of 22 items was used three times to measure perceptions of: (a) minimally-acceptable library performance on the SERVQUAL dimensions, (b) desired library performance on the SERVQUAL dimensions, and (c) perceived actual library performance on the SERVQUAL dimensions. Each item was rated using a "1" ("low") to "9" ("high") Likert-type response format.

Results

Reliability Analyses

The reliability of the SERVQUAL scores was evaluated by computing Cronbach's alpha coefficients across various partitions of the sample. These results are presented in Table 2. Alpha is a variance-accounted-for statistic that estimates the proportion of score variance that is systematic. However, mathematically the coefficient can be negative and even less than -1, under particularly dire measurement circumstances (see Reinhardt, 1996).

INSERT TABLE 2 ABOUT HERE.

Factor Analyses

Several principal components analyses were conducted, following the admonitions of Thompson and Daniel (1996), to evaluate SERVQUAL score validity. The first analysis focused on responses of the 697 participants to all 66 (22 items by 3 ratings frameworks--(a) minimally-acceptable library performance, (b) desired performance, and (c) perceived actual performance) items.

The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy for this analysis was .94, a clearly superior value due to our large sample size. Indeed, our sample size was considerably larger than those in any of the SERVQUAL library studies of various sorts cited by Nitecki (1997, p. 182).

Based on application of Cattell's visual "scree" test, three components accounting for 54.1% of the item covariance were extracted and rotated to the varimax criterion. The eigenvalues (λ) for the first three factors prior to rotation (Thompson, 1989) were 20.4, 8.7, and 6.6.

The three ratings frameworks (e.g., minimally-acceptable services) clearly emerged as the three factors in this analysis. Each item was "univocal" (i.e., was salient [pattern/structure coefficient > |.45|] to only one factor). Every one of the 66 items was salient to the perceptual framework that the item purportedly measured.

Next, the 22 items within each of the three measurement frameworks were analyzed separately to determine if the five SERVQUAL scales (i.e., *tangibles, reliability, responsiveness, assurance* and *empathy*) were recoverable.

The KMO statistic for the "minimum-expectation" ratings was .97. Based on a "scree" analysis, three principal components were extracted. Because a "simple structure" did not emerge after varimax rotation, the factors were then rotated to the promax criterion (Gorsuch, 1983). The factor pattern (i.e., weights analogous to regression beta weights) and structure coefficients (i.e., correlations between scores on the 22 items with scores on the 3 factors) from this analysis are presented in Table 3 (λ_1 = 12.9, λ_2 = 1.3, and λ_3 = .9).

INSERT TABLE 3 ABOUT HERE.

The KMO statistic for the "desired" ratings was .96. Table 4 presents the factor pattern and structure coefficients from a promax rotation of these results (λ_1 = 10.8, λ_2 = 1.6, and λ_3 = 1.0).

INSERT TABLE 4 ABOUT HERE.

The KMO statistic for the "perceived" ratings was .97. Table 5 presents the factor pattern and structure coefficients from a promax rotation of these results (λ_1 = 11.8, λ_2 = 1.2, and λ_3 = 1.0).

INSERT TABLE 5 ABOUT HERE.

Discussion

Score Reliability

The alpha coefficients reported in Table 2 were uniformly high across various scales, and across partitions of the sample by both years and user groups. SERVQUAL scores tended to be slightly less

reliable on the tangibles and the assurance scales and most reliable on the reliability scale. These results lend some credence to an expectation that SERVQUAL score quality tends to be fairly reasonable across both time and user group variations. Such is not always the case (Vacha-Haase, 1998).

Factor Analytic Results

Regarding the factor analysis results bearing upon the construct validity of SERVQUAL scores, when used in the context of evaluating library services, our results were less favorable. On the one hand, it is noteworthy that our factor analysis of the 66 items (pooled across the three frames of reference) did perfectly recover the three reference frames. Clearly, the 697 respondents attended to these reference frames and were readily able to distinguish them from each other. It is also noteworthy that an orthogonal rotation (i.e., varimax) recovered these three factors, meaning that an uncorrelated score structure reflecting the three frameworks was plausible.

On the other hand, however, the three separate analyses of the 22 SERVQUAL items computed independently within the three reference frames (i.e., (a) minimally-acceptable library performance, (b) desired performance, and (c) perceived actual performance) did not recover the five dimensions (i.e., *tangibles, reliability, responsiveness, assurance* and *empathy*) conventionally computed for SERVQUAL data. These results are consistent with previous factor analytic findings with the measure (cf. Nitecki, 1996b).

At most three factors underlay the three sets of responses to the 22 items. And even these factors were fairly highly correlated,

with factor correlations ranging from .474 (r^2 = 22.5%) to .640 (r^2 = 41.0%), as reported in Tables 3 through 5.

It is instructive to compare interpretations of the factors across the three frames of reference, because these comparisons make clear that score validity can vary across measurement contexts. Our results suggest that direct comparisons of scores on five dimensions across the three frames of reference would be very misleading.

A "Tangibles" factor emerged as the third factor in all three analyses. However, even its composition varied somewhat across the analyses, as regards the presence or absence of the item, "modern equipment."

Minimum-expectations Factors. The primary factor in this analysis appears to be "Service Efficacy"--the service experience is productive for users. As reported in Table 3, the underlying construct is particularly reflected in "providing services as promised" (Pattern coefficient = .714; r_s = .823), "employees have knowledge to answer customers' questions" (P = .668; r_s = .814), use of "modern equipment" (P = .827; r_s = .806), "convenient business hours" (P = .772; r_s = .789), and "maintaining error-free records" (P = .724; r_s = .781).

The second factor appears to involve "Affect of Service Experience"--patrons feel that service is caring and client-oriented. As reported in Table 3, the underlying construct is particularly reflected in "employees who are consistently courteous" (Pattern coefficient = .866; r_s = .864), "employees who deal with customers in a caring fashion" (P = .790; r_s = .847), and

"willingness to help customers" (P = .620; r_s = .823).

Desired Factors. The primary factor in this analysis appears to be "Staff Service Orientation"--customers perceive staff to be service-oriented. As reported in Table 4, the underlying construct is particularly reflected in "willingness to help customers" (Pattern coefficient = .757, r_s = .813), "providing service at the promised time" (P = .770; r_s = .777), "employees who are consistently courteous" (P = .734; r_s = .780), "having the customers' best interests at heart" (P = .733; r_s = .780), and "dependability in handling customers' service problems" (P = .622; r_s = .782).

The second factor appears to involve "Service Efficiency" --patrons feel that service is efficiently provided. As reported in Table 4, the underlying construct is particularly reflected in "modern equipment" (Pattern coefficient = .781; r_s = .802), "convenient business hours" (P = .721; r_s = .742), "performing services right the first time" (P = .536; r_s = .716), and "employees have knowledge to answer customers' questions" (P = .497; r_s = .716).

Perceived Factors. The primary factor appears to involve "Affect of Service Experience"--patrons feel that service is caring and client-oriented. As reported in Table 5, the underlying construct is particularly reflected in "employees who are consistently courteous" (Pattern coefficient = .977; r_s = .865), "employees who deal with customers in a caring fashion" (P = .907; r_s = .877), "willingness to help customers" (P = .712; r_s = .847), and "having the customers' best interests at heart" (P = .609; r_s

=. .804).

The second factor appears to involve "Service Reliability" --patrons feel that service is reliably provided. As reported in Table 5, the underlying construct is particularly reflected in "providing services as promised" (Pattern coefficient = .902; r_s = .882), "performing services right the first time" (P = .690; r_s = .810), "keeping customers informed when services are to be performed" (P = .697; r_s = .786), "dependability in handling customers' service problems" (P = .571; r_s = .788), and "performing services at the promised time" (P = .689; r_s = .781).

Implications for Library Service Evaluations

One implication of these results is that users appear to employ frameworks for thinking about library services that reflect subtle but important differences when processing questions about (a) minimally-acceptable library performance, (b) desired performance and (c) perceived, actual performance. In that these differences exist, the underlying theoretical framework of the SERVQUAL gap model is brought into question. Parasuraman, Berry and Zeithaml (1988, 1991) operationalized a service quality construct using a discrepancy model that compares customer perceptions of service against expected service. Parasuraman et al. (1994) defined two measures to analyze the difference between expectations and perceptions: *MSS*, the measure of service superiority, the difference between desired and perceived service, and *MSA*, the measure of service adequacy, the difference between perceived and minimum service. They reported the results of a factor analysis of MSS and MSA scores in which the factor structures were similar,

thus implying that ready comparisons could be made between MSS and MSA discrepancy scores.

But our results suggest that although respondents can readily discern the differences among minimum, desired and perceived response frameworks, the underlying factor structures are not the same. This possibility was recognized by Babakus and Boller (1992), who noted "empirical evidence suggests that difference scores such as these typically have unstable factor structures from one application to another" (p. 256). Others corroborate their findings (Andaleeb & Simmonds, 1998; Brown, Churchill & Peter, 1993; Van Dyke, Kappelman & Prybutok, 1997). Van Dyke et al. suggested then that "The direct measurement of one's perception of service quality that is the outcome of this cognitive evaluation process seems more likely to yield a valid and reliable outcome. If the discrepancy is what one wants to measure, then one should measure it directly" (p. 197). Cronin and Taylor (1992) found that the perceptions component of perception/expectation scores consistently performed as a better predictor of service quality than the difference score itself. As a result, Andaleeb and Simmonds (1998) abandoned the discrepancy score method in their study and gathered perceptions data only.

Conclusions

As a premise in designing SERVQUAL, Parasuraman, Zeithaml and Berry (1985) posited the existence of a second-order abstraction of service quality that is conceptually generalizable across industries, brands and product classes, and hence a measurement scale that permits cross-industry and cross-product comparisons. This model presumed five interrelated, first-order dimensions. Much

of the criticism of SERVQUAL to date has centered on the definition of these five dimensions.

Authors of several studies have concluded that the SERVQUAL instrument does not consistently measure the same factors, and that indeed structure is context-specific (e.g., Babakus & Boller, 1992; Carman, 1990; Van Dyke et al., 1997). Few seem to argue that SERVQUAL measures quality to some extent; however, the underlying factors defining quality seem to be partially inconsistent across service providers or contexts. Our results, as well as others reported in the literature on applications in academic libraries (Andaleeb & Simmonds, 1998; Nitecki, 1996a), lend credence to this view.

Nitecki (1996a) reported the results of a factor analysis yielding three rather than five factors. Supporting the results of our analysis, Nitecki reported that only the Tangibles dimension emerged as a discrete recognizable factor across all three of her analyses of ILL Reserve and Reference services. However, there is a noteworthy similarity in Nitecki's combined results examining three services and those in our study of library services at Texas A&M University for 1995, 1997 and 1999. Our factor analysis for Perceived scores very closely replicates Nitecki's results with the exception of three items: (a) "convenient business hours" correlated with Tangibles in our assessment rather than Nitecki's Factor 1, (b) "dependability in handling customers' service problems" correlated with our Factor II rather than Nitecki's Factor 1, and (c) "assuring customers of the accuracy and confidentiality of transactions" correlated with our Factor II

rather than Nitecki's Factor 1.

While the five dimensions in Parasuraman et al.'s model have not been recovered in studies conducted in academic library settings, three factors seem to have emerged consistently. Andaleeb and Simmonds' (1998) factor analysis of an alternative set of constructs, based loosely on SERVQUAL dimensions, isolated a factor, *Demeanor,* constituting one of the two most important factors underlying quality in library service. Demeanor is a rough combination of two SERVQUAL dimensions, empathy and assurance, and the helpfulness criterion normally associated with responsiveness. It is possible that the Demeanor factor in the Andaleeb and Simmonds study speaks to a similar concept in our and Nitecki's studies identified in the factor, "Effect of Service Experience," which is a close, but not exact amalgamation of SERVQUAL's responsiveness, assurance and empathy dimensions. Parasuraman et al. suggested in 1994 that there may be some overlap among responsiveness, assurance and empathy dimensions, and that these elements may constitute one rather than three factors. Other studies in retailing and in banking, motor vehicle, brokerage, electrical appliance and life insurance services industries have yielded similar results (Dabholkar, Thorpe & Rentz, 1996).

Library managers are well advised to exercise caution in interpreting results of SERVQUAL studies based upon the five-dimension model. While our results indicate consistent score reliability across the three years in which the studies were conducted, it is important for researchers to remember that reliability is score-specific and not instrument-specific and may

vary across each administration (Vacha-Haase, 1998). Our results also indicate that different factor structures underlie responses to minimum, desired and perceived responses, and so the practice of calculating difference scores by subtraction across these frameworks may be dubious. The use of perceived scores alone would also simplify the instrument considerably.

It is widely understood that new measures are needed to judge the quality of services and collections in research libraries. In recognition of this need the ARL Board recently appointed a New Measures Group to frame the questions for a discussion of new measures. There is wide agreement that user satisfaction is one of the key factors in assessing whether research libraries satisfy their missions to host institutions and to society at large. Franklin and Nitecki (1999) in their white paper on user satisfaction stated the problems: (1) "The primary issues at this juncture are whether a more standardized approach to assessing user satisfaction and service quality can be developed and, if so, whether such a standardized approach might yield comparable data that would be useful to ARL libraries" and (2) "Could a standard set of assessment variables be developed and then offered for application at several libraries? Can user satisfaction and user-based judgements of library service quality contribute to our understanding of library impact or value?" (p. 6).

One of the underlying questions in our study was whether SERVQUAL can be applied generally in research libraries as well as strategically at individual libraries. This study, in concert with those of Nitecki (1996) and Andaleeb and Simmonds (1998),

represents one step in devising a tool for assessing quality library service capable of wide application. As a whole SERVQUAL seems to measure quality in libraries as a higher-order concept that holds some promise of reasonably universal application in academic libraries. However, studies to date indicate fairly consistently that there are three rather than five factors underlying perceptions of quality service in academic libraries.

Whether an adapted SERVQUAL can answer the challenge for a standardized protocol to compare libraries remains to be seen. While acknowledging the wisdom of Hernon and Calvert's (1996) exhortation to the unwary that, "It is not possible to develop a generic instrument applicable to all libraries in all circumstances" (p. 388), the need to understand what constitutes quality service for library users is undeniable, for "we cannot manage what we cannot measure" (Van Dyke et al., 1997, p. 205). Libraries must be responsive to user expectations, and in order to do so, we must better understand how users judge quality in library services.

References

Andaleeb, S.S., & Simmonds, P.L. (1998). Explaining user satisfaction with academic libraries: Strategic implications. College and Research Libraries, 59, 156-167.

Babakus, E., & Boller, G.W. (1992). An empirical assessment of the SERVQUAL scale. Journal of Business Research, 24, 253-268.

Brown, T.J., Churchill, G.A., & Peter, J.P. (1993). Improving the measurement of service quality. Journal of Retailing, 66, 127-139.

Carman, J.M. (1990). Consumer perceptions of service quality: An assessment of SERVQUAL dimensions. Journal of Retailing, 66, 33-55.

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. New York: Holt, Rinehart and Winston.

Cronin, J.J., & Taylor, S.A. (1992). Measuring service quality: A reexamination and extension. Journal of Marketing, 56, 55-68.

Dabholkar, P.A., Thorpe, I.D., & Rentz, J.O. (1996). A measure of service quality for retail stores: Scale development and validation. Journal of the Academy of Marketing, 24, 3-16.

Franklin, B., & Nitecki, D. (1999). User satisfaction white paper. Washington, DC: Association of Research Libraries.

Gorsuch, R.L. (1983). Factor analysis (2nd ed.). Hillsdale, NJ: Erlbaum.

Gronlund, N.E., & Linn, R.L. (1990). Measurement and evaluation in teaching (6th ed.). New York: Macmillan.

Hernon, P. & Calvert, P. (1996). Methods for measuring service quality in university libraries in New Zealand. Journal of

Academic Librarianship, 22, 387-391.

Hernon, P., & McClure, C.R. (1990). Evaluation and library decision making. Norwood, NJ: Ablex.

Nitecki, D.A. (1996a). An assessment of the applicability of SERVQUAL dimensions a customer-based criteria for evaluating quality of services in an academic library (Doctoral dissertation, University of Maryland, 1995). Dissertation Abstracts International, 56, 2918A. (University Microfilms No. 95-39,711)

Nitecki, D.A. (1996b). Changing the concept and measure of service quality in academic libraries. The Journal of Academic Librarianship, 22, 181-190.

Nitecki, D.A. (1997). Assessment of service quality in academic libraries: Focus on the applicability of the SERVQUAL. Proceedings of the second Northumbria International Conference on Performance Measurement in Libraries and Information Services (pp. 181-196). Newcastle upon Tyne, England: University of Northumbria at Newscastle.

Parasuraman, A., Berry, L.L., & Zeithaml, V.A. (1988). SERVQUAL: A multiple-item scale for measuring customer perceptions of service quality. Journal of Retailing, 64, 12-40.

Parasuraman, A., Berry, L.L. & Zeithaml, V.A. (1991). Refinement and reassessment of the SERVQUAL scale. Journal of Retailing, 67, 420-450.

Parasuraman, A., Zeithaml, V.A., & Berry, L.L. (1985). A conceptual model of service quality and its implications for future research. Journal of Marketing, 70, 41-50.

Parasuraman, A., Zeithaml, V.A., & Berry, L.L. (1994). Alternative scales for measuring service quality: A comparative assessment based on psychometric and diagnostic criteria. Journal of Retailing, 49, 201-230.

Reinhardt, B. (1996). Factors affecting coefficient alpha: A mini Monte Carlo study. In B. Thompson (Ed.), Advances in social science methodology (Vol. 4, pp. 3-20). Greenwich, CT: JAI Press.

Rowley, G.L. (1976). The reliability of observational measures. American Educational Research Journal, 13, 51-59.

Stein, J. (1997). Feedback from a captive audience: Reflections on the results of a SERVQUAL survey of interlibrary loan services at Carnegie Mellon University libraries. Proceedings of the second Northumbria International Conference on Performance Measurement in Libraries and Information Services (pp. 207-222). Newcastle upon Tyne, England: University of Northumbria at Newscastle.

Thompson, B. (1989). Prerotation and postrotation eigenvalues shouldn't be confused: A reminder. Measurement and Evaluation in Counseling and Development, 22, 114-116.

Thompson, B. (1994). Guidelines for authors. Educational and Psychological Measurement, 54, 837-847.

Thompson, B., & Daniel, L.G. (1996). Factor analytic evidence for the construct validity of scores: An historical overview and some guidelines. Educational and Psychological Measurement, 56, 213-224.

Vacha-Haase, T. (1998). Reliability generalization: Exploring

variance in measurement error affecting score reliability across studies. Educational and Psychological Measurement, 58, 6-20.

Van Dyke, T.P., Kappelman, L.A., & Prybutok, V.R. (1997). Measuring information systems service quality: Concerns on the use of the SERVQUAL questionnaire. MIS Quarterly, 21, 195-208.

Zeithaml, V.A., Parasuraman, A., Berry, L.L. (1990). Delivering quality service: Balancing customer perceptions and expectations. New York: Free Press.

Table 1
Participants Broken Down by
Role Group and Year

Role Group	Year 1995	1997	1999	Row Total
staff	23	52	26	101 (14.5)
undergrad	67	65	37	169 (24.2)
faculty	32	78	60	170 (24.4)
graduate	57	92	108	257 (36.9)
Column Total	179 (25.7)	287 (41.2)	231 (33.1)	697 (100.0)

Note. Percentages are presented within parentheses.

Table 2
Alpha Coefficients Across Sample Partitions

Sample (n)/ Scale	Referent		
	Minimum	Desired	Perceived
Time			
1995 (n = 179)			
Tangibles	0.816	0.739	0.796
Reliability	0.881	0.814	0.890
Responsiveness	0.862	0.768	0.847
Assurance	0.850	0.762	0.837
Empathy	0.871	0.765	0.844
1997 (n = 287)			
Tangibles	0.853	0.746	0.786
Reliability	0.899	0.871	0.881
Responsiveness	0.878	0.868	0.875
Assurance	0.821	0.766	0.822
Empathy	0.885	0.835	0.829
1999 (n = 231)			
Tangibles	0.781	0.758	0.777
Reliability	0.909	0.870	0.864
Responsiveness	0.872	0.803	0.843
Assurance	0.803	0.726	0.776
Empathy	0.884	0.834	0.856
Role Group			
Faculty (n = 170)			
Tangibles	0.796	0.734	0.806
Reliability	0.897	0.849	0.902
Responsiveness	0.840	0.812	0.877
Assurance	0.804	0.745	0.823
Empathy	0.871	0.800	0.883
Staff (n = 101)			
Tangibles	0.859	0.718	0.736
Reliability	0.865	0.855	0.883
Responsiveness	0.868	0.858	0.877
Assurance	0.838	0.794	0.816
Empathy	0.889	0.822	0.868
Undergraduate students (n = 169)			
Tangibles	0.840	0.761	0.793
Reliability	0.917	0.850	0.882
Responsiveness	0.894	0.815	0.847
Assurance	0.865	0.781	0.787
Empathy	0.900	0.859	0.821
Graduate students (n = 257)			

Tangibles	0.808	0.766	0.781
Reliability	0.893	0.872	0.853
Responsiveness	0.861	0.837	0.843
Assurance	0.788	0.721	0.811
Empathy	0.867	0.806	0.814
Total (n = 697)			
Tangibles	0.822	0.749	0.785
Reliability	0.899	0.860	0.878
Responsiveness	0.871	0.828	0.858
Assurance	0.823	0.751	0.810
Empathy	0.882	0.821	0.842

Table 3
Promax-rotated Pattern and Structure Matrices
and Factor Correlation Matrix for *Minimum-expectation* Items (n = 697)

Item	Pattern			Structure		
	I	II	III	I	II	III
21("T") Modern equipment	.827	.249	.275	.806	.414	.555
20("E") Convenient business hours	.772	.122	.189	.789	.462	.508
9("L") Maintaining error-free customer and catalog records	.724	.123	.004	.781	.557	.380
11("L") Providing services as promised	.714	.279	.133	.823	.659	.364
13("A") Employees have knowledge to answer customers' questions	.668	.233	.000	.814	.653	.451
16("L") Performing services right the first time	.635	.240	.000	.786	.640	.438
14("L") Dependability in handling customers' service problems	.560	.333	.008	.811	.730	.534
15("R") Readiness to respond to customers' questions	.522	.386	.008	.803	.754	.536
10("R") Keeping customers informed when services be performed	.446	.311	.169	.726	.680	.551
22("A") Assure customers accuracy/confident of transactions	.430	.009	.358	.666	.550	.620
2("A") Employees who are consistently courteous	.005	.866	.005	.524	.864	.479
3("E") Employees who deal with customers in a caring fashion	.147	.790	.289	.493	.847	.626
7("E") Having the customers' best interest at heart	.133	.664	.158	.629	.830	.568
8("R") Willingness to help customers	.367	.620	.005	.731	.823	.450
5("E") Employees who understand the needs of their customers	.280	.584	.007	.684	.799	.515
4("L") Providing service at the promised time	.515	.516	.207	.737	.733	.315
1("R") Prompt service to customers	.336	.462	.005	.652	.700	.457
12("A") Employees who instill confidence in customers	.005	.449	.445	.551	.710	.701
18("E") Giving customers individual attention	.295	.386	.300	.687	.727	.646
17("T") Visually appealing facilities	.006	.002	.852	.469	.458	.871
19("T") Employees who have a neat, professional appearance	.008	.156	.827	.421	.530	.865
6("T") Visually appealing materials associated with the service	.216	.114	.568	.568	.544	.734

Note. "T" = tangibles; "L" = reliability; "R" = responsiveness; "A" = assurance; "E" = empathy. The pivot power used in the promax rotation was 3. The factor correlation coefficients were: $r_{I \times II} = .629$; $r_{I \times III} = .494$; $r_{II \times III} = .518$. Factor pattern coefficients greater than |.4| are underlined.

Table 4
Promax-rotated Pattern and Structure Matrices
and Factor Correlation Matrix for *Desired* Items (n = 697)

Item		Pattern			Structure		
		I	II	III	I	II	III
4("L")	Providing service at the promised time	.770	.190	.199	.777	.539	.225
8("R")	Willingness to help customers	.757	.116	.001	.813	.516	.381
2("A")	Employees who are consistently courteous	.734	.002	.118	.780	.410	.460
7("E")	Having the customers' best interest at heart	.733	.133	.248	.780	.336	.555
3("E")	Employees who deal with customers in a caring fashion	.648	.218	.394	.719	.251	.634
11("L")	Providing services as promised	.626	.312	.114	.738	.610	.279
14("L")	Dependability in handling customers' service problems	.622	.295	.001	.782	.629	.393
15("R")	Readiness to respond to customers' questions	.620	.291	.003	.791	.632	.418
1("R")	Prompt service to customers	.584	.158	.006	.696	.488	.384
5("E")	Employees who understand the needs of their customers	.581	.203	.009	.731	.540	.425
12("A")	Employees who instill confidence in customers	.543	.168	.436	.660	.258	.641
10("R")	Keeping customers informed when services be performed	.436	.321	.162	.685	.605	.469
21("T")	Modern equipment	.005	.781	.155	.439	.802	.374
20("E")	Convenient business hours	.004	.721	.129	.410	.742	.336
16("L")	Performing services right the first time	.334	.536	.000	.622	.716	.329
13("A")	Employees have knowledge to answer customers' questions	.467	.497	.010	.686	.716	.278
9("L")	Maintaining error-free customer and catalog records	.382	.429	.002	.602	.627	.296
19("T")	Employees who have a neat, professional appearance	.000	.001	.856	.401	.257	.853
17("T")	Visually appealing facilities	.151	.187	.829	.342	.364	.816
6("T")	Visually appealing materials associated with the service	.008	.194	.576	.456	.416	.674
22("A")	Assure customers accuracy/confident of transactions	.247	.223	.367	.541	.470	.554
18("E")	Giving customers individual attention	.323	.224	.387	.627	.518	.610

Note. "T" = tangibles; "L" = reliability; "R" = responsiveness; "A" = assurance; "E" = empathy. The pivot power used in the promax rotation was 3. The factor correlation coefficients were: $r_{I \times II} = .534$; $r_{I \times III} = .474$; $r_{II \times III} = .534$. Factor pattern coefficients greater than |.4| are underlined.

Table 5
Promax-rotated Pattern and Structure Matrices
and Factor Correlation Matrix for *Perceived* Items (n = 697)

Item		Pattern			Structure		
		I	II	III	I	II	III
2("A")	Employees who are consistently courteous	.977	.167	.000	.865	.453	.425
3("E")	Employees who deal with customers in a caring fashion	.907	.005	.001	.877	.531	.467
8("R")	Willingness to help customers	.712	.220	.001	.847	.670	.489
1("R")	Prompt service to customers	.613	.263	.002	.770	.644	.448
7("E")	Having the customers' best interest at heart	.609	.229	.009	.804	.667	.541
18("E")	Giving customers individual attention	.546	.183	.199	.769	.638	.589
5("E")	Employees who understand the needs of their customers	.531	.355	.000	.763	.700	.484
15("R")	Readiness to respond to customers' questions	.531	.296	.010	.773	.688	.542
12("A")	Employees who instill confidence in customers	.459	.319	.109	.722	.671	.526
13("A")	Employees have knowledge to answer customers' questions	.431	.425	.002	.713	.711	.479
11("L")	Providing services as promised	.001	.902	.002	.553	.882	.451
10("R")	Keeping customers informed when services be performed	.010	.697	.005	.570	.786	.477
16("L")	Performing services right the first time	.169	.690	.002	.623	.810	.483
4("L")	Providing service at the promised time	.244	.689	.121	.620	.781	.379
9("L")	Maintaining error-free customer and catalog records	.000	.662	.132	.496	.734	.487
14("L")	Dependability in handling customers' service problems	.302	.571	.004	.692	.788	.513
22("A")	Assure customers accuracy/confident of transactions	.005	.412	.366	.510	.640	.614
17("T")	Visually appealing facilities	.010	.136	.817	.450	.365	.798
21("T")	Modern equipment	.003	.010	.744	.432	.478	.781
19("T")	Employees who have a neat, professional appearance	.224	.003	.643	.548	.455	.746
6("T")	Visually appealing materials associated with the service	.235	.100	.518	.577	.528	.698
20("E")	Convenient business hours	.256	.421	.492	.277	.521	.580

Note. "T" = tangibles; "L" = reliability; "R" = responsiveness; "A" = assurance; "E" = empathy. The pivot power used in the promax rotation was 3. The factor correlation coefficients were: $r_{I \times II} = .640$; $r_{I \times III} = .536$; $r_{II \times III} = .535$. Factor pattern coefficients greater than |.4| are underlined.

Table A.1
Varimax-rotated Pattern/Structure Coefficient Matrix
Across All Items (n = 697)

Item	Factor I	II	III
QM6	.663	.189	.126
QM17	.614	.159	.112
QM19	.598	.267	.127
QM21	.708	.026	.142
QM4	.744	.077	.173
QM9	.710	.025	.171
QM11	.767	.072	.173
QM14	.821	.111	.202
QM16	.758	.047	.190
QM1	.722	.125	.134
QM8	.802	.138	.180
QM10	.755	.144	.193
QM15	.825	.122	.210
QM2	.712	.180	.185
QM12	.712	.172	.174
QM13	.769	.093	.202
QM22	.688	.158	.161
QM3	.733	.133	.130
QM5	.786	.137	.167
QM7	.768	.162	.215
QM18	.783	.179	.153
QM20	.716	.066	.082
QP6	.164	.649	.111
QP17	.110	.543	.140
QP19	.138	.617	.140
QP21	.026	.587	.147
QP4	.149	.721	.056
QP9	.080	.676	.072
QP11	.106	.748	.101
QP14	.087	.802	.052
QP16	.098	.761	.099
QP1	.183	.743	.061
QP8	.115	.815	.039
QP10	.141	.714	.121
QP15	.110	.789	.133
QP2	.119	.705	.102

QP12	.086	.770	.097
QP13	.043	.781	.056
QP22	.139	.651	.088
QP3	.133	.765	.067
QP5	.099	.790	.068
QP7	.129	.801	.091
QP18	.163	.772	.063
QP20	.106	.480	.114
QD6	.128	.124	.565
QD17	.122	.148	.506
QD19	.133	.228	.505
QD21	.116	.010	.632
QD4	.145	.035	.717
QD9	.094	-.009	.668
QD11	.136	.062	.728
QD14	.146	.083	.787
QD16	.172	.049	.694
QD1	.166	.078	.674
QD8	.155	.114	.759
QD10	.176	.057	.729
QD15	.186	.086	.791
QD2	.174	.098	.719
QD12	.150	.117	.635
QD13	.105	.066	.736
QD22	.196	.091	.601
QD3	.140	.166	.664
QD5	.164	.111	.721
QD7	.131	.140	.721
QD18	.182	.148	.681
QD20	.110	.085	.572

Note. Coefficients greater than |.45| are underlined.

Table A.2
Varimax-rotated Pattern/Structure Coefficient Matrix
for *Minimum-expectation* Items (n = 697)

	Factor		
Item	I	II	III
QM6	.360	.286	.625
QM17	.234	.178	.821
QM19	.150	.292	.809
QM21	.743	.072	.397
QM4	.600	.573	.067
QM9	.701	.312	.171
QM11	.723	.430	.122
QM14	.645	.482	.307
QM16	.666	.401	.220
QM1	.473	.526	.255
QM8	.531	.656	.206
QM10	.551	.447	.354
QM15	.626	.517	.305
QM2	.250	.783	.272
QM12	.288	.515	.556
QM13	.694	.404	.226
QM22	.506	.285	.472
QM3	.184	.735	.452
QM5	.466	.626	.296
QM7	.374	.674	.362
QM18	.464	.498	.462
QM20	.715	.152	.337

Note. Coefficients greater than |.35| are underlined.

Table A.3
Varimax-rotated Pattern/Structure Coefficient Matrix
for *Desirability* Items (n = 697)

Item	Factor I	Factor II	Factor III
QD6	.231	.284	**.609**
QD17	.066	.244	**.793**
QD19	.164	.102	**.831**
QD21	.170	**.755**	.249
QD4	**.716**	**.376**	.038
QD9	**.451**	**.519**	.144
QD11	**.629**	**.463**	.100
QD14	**.645**	**.462**	.214
QD16	**.437**	**.611**	.169
QD1	**.587**	.326	.235
QD8	**.722**	.326	.205
QD10	**.510**	**.454**	.320
QD15	**.647**	**.461**	.239
QD2	**.694**	.208	.308
QD12	**.544**	.051	**.544**
QD13	**.531**	**.597**	.100
QD22	**.352**	.333	**.455**
QD3	**.621**	.027	**.524**
QD5	**.601**	**.372**	.268
QD7	**.691**	.114	**.418**
QD18	**.426**	**.358**	**.495**
QD20	.166	**.699**	.219

Note. Coefficients greater than |.35| are underlined.

Table A.4
Varimax-rotated Pattern/Structure Coefficient Matrix
for *Perceived* Items (n = 697)

	Factor		
Item	I	II	III
QP6	.371	.282	.577
QP17	.244	.093	.761
QP19	.349	.186	.656
QP21	.183	.250	.721
QP4	.413	.675	.133
QP9	.243	.633	.301
QP11	.270	.812	.213
QP14	.468	.623	.273
QP16	.379	.685	.250
QP1	.639	.423	.210
QP8	.719	.417	.234
QP10	.321	.676	.261
QP15	.604	.455	.310
QP2	.838	.146	.202
QP12	.548	.456	.307
QP13	.534	.524	.244
QP22	.264	.474	.468
QP3	.815	.232	.232
QP5	.601	.488	.242
QP7	.652	.416	.306
QP18	.606	.380	.378
QP20	.017	.417	.509

Note. Coefficients greater than |.35| are underlined.